

# Genetic architecture of seed glycerolipids in Asian cultivated rice

Jun Hong<sup>1,2</sup> | Leah Rosental<sup>3</sup> | Yang Xu<sup>4</sup> | Dawei Xu<sup>1</sup> | Isabel Orf<sup>3</sup> |  
Wengsheng Wang<sup>5</sup> | Zhiqiang Hu<sup>1</sup> | Su Su<sup>1</sup> | Shaoxing Bai<sup>1</sup> |  
Mohammed Ashraf<sup>1</sup>  | Chaoyang Hu<sup>1</sup> | Changquan Zhang<sup>6</sup> | Zhikang Li<sup>5</sup> |  
Jianlong Xu<sup>5</sup> | Qiaoquan Liu<sup>6</sup> | Hui Zhang<sup>7</sup> | Fengli Zhang<sup>1</sup> | Zhijing Luo<sup>1</sup> |  
Mingjiao Chen<sup>1</sup> | Xiaofei Chen<sup>1</sup> | Natalie Betts<sup>2</sup> | Alisdair Fernie<sup>8</sup> |  
Wanqi Liang<sup>1</sup> | Guanqun Chen<sup>4</sup>  | Yariv Brotman<sup>3</sup> | Dabing Zhang<sup>1,2</sup> |  
Jianxin Shi<sup>1</sup> 

<sup>1</sup>Department of Genetics and Developmental Science, Joint International Research Laboratory of Metabolic and Developmental Sciences, State Key Laboratory of Hybrid Rice, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China

<sup>2</sup>Waite Research Institute, School of Agriculture, Food and Wine, University of Adelaide, Urrbrae, SA, Australia

<sup>3</sup>Department of Life Sciences, Ben Gurion University of the Negev, Beersheva, Israel

<sup>4</sup>Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada

<sup>5</sup>Department of Rice Molecular Design Technology and Application, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, China

<sup>6</sup>Department of Agronomy, Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding, College of Agriculture, Yangzhou University, Yangzhou, China

<sup>7</sup>Department of Plant Science, School of Life Sciences, Shanghai Normal University, Shanghai, China

<sup>8</sup>Department of Central Metabolism, Max Planck Institute of Molecular Plant Physiology, Potsdam-Golm, Germany

## Correspondence

Jianxin Shi and Dabing Zhang, Department of Genetics and Developmental Science, Joint International Research Laboratory of Metabolic and Developmental Sciences, State Key Laboratory of Hybrid Rice, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai, China.  
Email: [Jianxin.shi@sjtu.edu.cn](mailto:Jianxin.shi@sjtu.edu.cn) and [zhangdb@sjtu.edu.cn](mailto:zhangdb@sjtu.edu.cn)

Guanqun Chen, Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada.  
Email: [gc24@ualberta.ca](mailto:gc24@ualberta.ca)

Yariv Brotman, Department of Life Sciences, Ben Gurion University of the Negev, Beersheva, Israel.  
Email: [brotmany@post.bgu.ac.il](mailto:brotmany@post.bgu.ac.il)

## Funding information

National Natural Sciences Foundation of China

## Abstract

Glycerolipids are essential for rice development and grain quality but its genetic regulation remains unknown. Here we report its genetic base using metabolite-based genome-wide association study and metabolite-based quantitative traits locus (QTL) analyses based on lipidomic profiles of seeds from 587 Asian cultivated rice accessions and 103 chromosomal segment substitution lines, respectively. We found that two genes encoding phosphatidylcholine (PC):diacylglycerol cholinephosphotransferase (*OsLP1*) and granule-bound starch synthase I (*Waxy*) contribute to variations in saturated triacylglycerol (TAG) and lyso-PC contents, respectively. We demonstrated that allelic variation in *OsLP1* sequence between *indica* and *japonica* results in different enzymatic preference for substrate PC-16:0/16:0 and different saturated TAG levels. Further evidence demonstrated that *OsLP1* also affects heading date, and that co-selection of *OsLP1* and a flooding-tolerant QTL in *Aus* results in the abundance of saturated TAGs associated with flooding tolerance. Moreover, we revealed that the sequence polymorphisms in *Waxy* has pleiotropic effects on lyso-PC and amylose content. We proposed that rice seed glycerolipids

have been unintentionally shaped during natural and artificial selection for adaptive or import seed quality traits. Collectively, our findings provide valuable genetic resources for rice improvement and evolutionary insights into seed glycerolipid variations in rice.

#### KEYWORDS

lyso-phosphatidylcholine, natural variation, *Oryza sativa*, saturated triacylglycerol, seed lipidome

## 1 | INTRODUCTION

Domestication, a complex process of selection, is a good model for evolutionary studies (Izawa et al., 2009). Asian cultivated rice (*Oryza sativa* L.), which originated from its ancestral progenitor *Oryza rufipogon*, is dominated by *indica* (Xian) and *japonica* (Geng) subspecies and grown globally, from tropical to temperate zones, and from lowland ecosystems (paddy, deep water, or flood conditions) to upland rainfed areas (Gutaker et al., 2020). During the domestication of Asian cultivated rice, artificial selection or local adaptation have associated closely with sequence variation and gene functional diversification (Huang et al., 2012). For example, grain quality genes such as *Waxy* (Zhang et al., 2019), *qSW5/GW5* (QTL for seed width on chromosome 5; Shomura et al., 2008) and *Betaine aldehyde dehydrogenase2* (*BAD2*; Bradbury et al., 2005) have been shown to be targets of artificial selection. In addition, several adaptive QTL/genes, such as *Heading date 1* (*Hd1*; Kim et al., 2018) and *Heading date 4/Grain number, plant height and heading date 7* (*Hd4/Ghd7*; Xue et al., 2008) for photoperiod adaptation, *basic leucine zipper 73* (*bZIP73*) for temperature acclimation (Liu et al., 2018), *Teosinte branched 2* (*OsTB2*) for upland environments (Lyu et al., 2020), and *Semi-dwarf 1* (*SD1*) for periodic flooding resistance (Kuroha et al., 2018), have also been identified.

Beyond phenotypic and physiological changes, significant metabolic variations have also been uncovered between *indica* and *japonica* subspecies (Chen et al., 2014; Hu et al., 2014). For example, several primary metabolites including amino acids and lipids (Hu et al., 2014) and secondary metabolites including flavonoids, phenolamides and arabidopyl alcohol derivatives (Chen et al., 2014) can be used as biomarkers to distinguish *indica* from *japonica* varieties. Currently, integrative approaches, including genomics, metabolomics, quantitative genetics, metabolite-based genome-wide association analysis (mGWAS) and metabolite-based quantitative traits locus (mQTL), have been used extensively to dissect genetic and biochemical bases of metabolic diversity in domesticated rice (Chen et al., 2014, 2016; Gong et al., 2013; Matsuda et al., 2012). Emerging knowledge from these studies have made it possible to improve grain chemical composition beyond our traditional nutrition targets, such as oil and proteins (Chen et al., 2016; Wen et al., 2016), facilitating design of de novo domestication strategies for the biofortification of crop grains at the metabolite level (Fernie & Yan, 2019). However, although achievements have been made in identifying candidate genes and loci controlling several rice metabolic

traits (Chen et al., 2014, 2016; Gong et al., 2013; Peng et al., 2017; Zhan et al., 2020), little have focused on how natural and artificial selection affects metabolic profiles during rice domestication and diversification (Chen et al., 2021).

Glycerolipids are structurally diverse metabolites, consisting of a glycerol backbone with various combinations of fatty acids and head groups (Horn & Benning, 2016). Glycerolipids are the largest family of plant lipids with essential conserved functions in cell membrane dynamics and cellular communication, and are indispensable nutrients for humans and animals (Horn & Benning, 2016). Increasing evidence has demonstrated roles for plant glycerolipids in photosynthesis, signal transduction, vesicle trafficking, cytoskeletal rearrangement, as well as several developmental processes (Colin & Jaillais, 2020; Welti & Wang, 2004). In *Arabidopsis thaliana*, phosphatidylcholine (PC) binds to the core flowering regulator Flowering Timing (FT) protein to regulate flowering (Nakamura et al., 2014, 2019). In addition, FT is sequestered in *Arabidopsis* cellular membrane via binding to phosphatidylglycerol (PG), modulating the plants' response to temperature changes (Susila et al., 2021). In maize, high PhosphatidylCholine 1 introgression from teosinte underlies a large mQTL that modulates PC levels and induces early flowering (Barnes et al., 2021). Results of those studies link endogenous glycerolipids with flowering time regulation in plants, highlighting the key roles of glycerolipids in plant domestication. However, the function and genetic regulation of glycerolipid identity and abundance in rice remains elusive, largely due to the complicated nature of glycerolipid metabolism and complex interactions with other pathways (Lavell & Benning, 2019; Li et al., 2019), which severely hinders our understanding of how glycerolipids affect rice development, grain quality and adaptation.

Advances in high-throughput lipidomics have enabled the analysis of large number of lipids simultaneously (Hummel et al., 2011); here we have integrated lipidomics and quantitative genetics to generate a new understanding of the genetic base of glycerolipid metabolism in different rice subspecies. We have profiled more than 100 seed glycerolipid molecules in 587 Asian cultivated rice accessions and 103 chromosome segment substitution lines (CSSLs) using mass spectrometry-based lipidomics. We have then investigated into the genetic and biochemical bases of natural variations in glycerolipids between *indica* and *japonica* subspecies using integrated mGWAS, mQTL, genetic and biochemical approaches, and revealed that *OsLP1* and *Waxy* contribute to variation in saturated triacylglycerol (TAG) and

lyso-PC contents, respectively. Moreover, we have explored the effect of domestication on this divergence using combined population genetic analysis and demonstrated that natural variation of rice seed glycerolipids has been unwittingly selected by natural selection for adaptive traits (such as flooding resistance and heading date) and by artificial selection for grain quality traits (such as mouthfeel). Our findings provide not only a snapshot of the genetic architecture of rice seed glycerolipids, producing new breeding resources for utilizing lipidomics in genetic improvement, but also an evolutionary insight into domestication-associated changes in seed glycerolipid profiles.

## 2 | MATERIAL AND METHODS

### 2.1 | Plant materials and growth conditions

A panel of 587 rice accessions (Supporting Information: Data Set 1) was obtained from the 3 K Rice Genome Project (Wang et al., 2018). Thirty-day-old seedlings of each accession were transplanted to the experimental farm in Sanya (31.03°N, 121.45°E) in the winter season of 2014–2015 under natural conditions. The experimental design was a randomized complete block design (including two rows of each accession and eight plants in each row). Mature rice seeds from three different plants for each line were collected, pooled, frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until analysis as described previously (Hu et al., 2014). One hundred and three CSSL lines were generated by crossing Nipponbare as recipient and 9311 as donor parent, followed by a series of backcrosses to Nipponbare with marker-assisted selection and subsequent self-crosses (Xu et al., 2017). CSSL lines were planted under natural conditions in the paddy field in Yangzhou (18.26°N, 108.56°E) during the summer season of 2012 and in Sanya during the winter season of 2012–2013. The experimental design of CSSL lines was the same as those for GWAS germplasm, whereas three replicates of each line were collected (each replicate was composed of pooled seeds from three different plants).

### 2.2 | Lipid extraction

Lipids were extracted from seeds (whole grain with husk) as previously described (Hummel et al., 2011). Briefly, seeds were ground into fine powder under liquid nitrogen and 50 mg of frozen powder was homogenized at maximum speed using a ball mill (MM 301, Retsch) for two 1 min periods. Lipids were extracted with 1 ml of precooled ( $-20^{\circ}\text{C}$ ) solvent mixture consisting of methanol/methyl-tert-butyl-ether (v/v, 1:3).  $0.1\ \mu\text{g}\ \text{ml}^{-1}$  of PE 34:0 and  $0.1\ \mu\text{g}\ \text{ml}^{-1}$  of PC 34:0 was added as internal standards (Sigma-Aldrich). Samples were incubated at  $4^{\circ}\text{C}$  for 10 min, followed by a sonication at room temperature for 10 min. After adding 500  $\mu\text{l}$  of 25% (v/v) aqueous methanol, samples were vortexed and centrifuged for 5 min at  $4^{\circ}\text{C}$  (14,000g). Next, 500  $\mu\text{l}$  of the upper lipophilic phase was collected, dried under vacuum, and stored at  $-80^{\circ}\text{C}$  until use.

### 2.3 | Glycerolipid profiling

Glycerolipid profiling was carried out as previously described using ultra performance liquid chromatography-Fourier transform-mass spectrometry (UPLC-FT-MS, Waters; Hummel et al., 2011). Briefly, dried lipophilic extracts were resuspended in 500  $\mu\text{l}$  of Buffer B (see below) and 1  $\mu\text{l}$  was used for injection. The mobile phases consisted of Buffer A (1% 1 M ammonium acetate and 0.1% acetic acid in water) and Buffer B (acetonitrile/isopropanol, v/v, 7:3, supplemented with 1 M ammonium acetate and 0.1% acetic acid) and the mobile phase flow rate was 400  $\mu\text{l}\ \text{min}^{-1}$ . The gradient separation was 1 min 45% A, followed by 3 min linear gradient from 45% A to 35% A, 8 min from 25% to 11% A, and 3 min from 11% to 1% A. After being washed for 3 min with 1% A, the column was re-equilibrated with 45% A for 4 min.

All the spectra were recorded with an Exactive mass spectrometer (Thermo Fisher) equipped with an electrospray ionization interface in both positive and negative modes altering between full-scan and all-ion fragmentation scan mode (mass range 100–1500 m/z, capillary voltage 3.0 kV, sheath gas flow value 60, auxiliary gas flow value 35), from minute 1 to minute 20 of the UPLC gradients. The resolution was 10,000 (10 scans  $\text{s}^{-1}$ ) and the Orbitrap loading time was up to 100 ms with a target value of 106 ions. The capillary temperature was  $150^{\circ}\text{C}$  and the temperature of the drying gas in the heated electrospray source was  $350^{\circ}\text{C}$ . The skimmer and the tube lens voltage were 25 and 130 V, respectively.

### 2.4 | Peak extraction, alignment, and annotation

REFINER MS 10.0 (GeneData, <http://www.genedata.com>) or Xcalibur (Version 3.1, Thermo Fisher) were used to automatically and manually, respectively, process chromatograms, peak detection, and integration, as described (Lapidot-Cohen et al., 2020). In the automated approach, molecular masses, retention times, and associated peak intensities were extracted from the raw files. Data processing included the separation of the full scan spectra from the all-ion fragmentation spectra, and the removal of chemical noise. Excel or Access (Microsoft) were used for further peak filtering on the manually extracted spectra or the aligned data matrices. Obtained features (m/z at a certain retention time) were queried against an in-house lipid database including about 200 lipid species for further annotation. MS/MS fragmentation using collision-induced dissociation mass spectra (25 eV collision energy) was used for further validation of representatives of different lipid classes (Supporting Information: Data Set 2).

### 2.5 | Statistical analysis

The values of the coefficient of variation (CV) were calculated independently for each metabolite (using  $\log_2$ -transformed metabolite trait data) using  $\sigma/\mu$ , where  $\sigma$  and  $\mu$  are the SD and mean of each

metabolite in GWAS population, respectively. The metabolite trait data ( $\log_2$ -transformed) of CSSL lines in two seasons was used to estimate the broad-sense heritability ( $H^2$ ) of lipid metabolites.  $H^2$  was calculated using the following equation:  $H^2 = \text{var}(G)/(\text{var}(G) + \text{var}(E))$  using one-way analysis of variance (ANOVA), where  $\text{var}(G)$  and  $\text{var}(E)$  are the variance derived from genetic and environmental effects, respectively. The variance components were estimated from the variance component analysis using R package. In one-way ANOVA, accessions served as the random effect and three biological replications served as the replication effect (Chen et al., 2014).

## 2.6 | Population structure analysis

Single-nucleotide polymorphism (SNP) genotype data were downloaded from the OryzaSNP database (<http://snp-seek.irri.org/download.zul>) and the 3K RG 404k CoreSNP Data set was used for population structure analysis. The Bayesian clustering software fastStructure (Raj et al., 2014) was used to calculate varying levels of  $K$  ( $K = 1-25$ ) and the command chooseK.py was used to identify the model complexity that maximized the marginal likelihood ( $K = 7$ ). All 587 accessions used in this study had a maximum subpopulation component value  $\geq 0.7$  (Supporting Information: Data Set 1). Principal component analysis (PCA) of genetic variance was conducted using GCTA software (Yang et al., 2011) and calculated using SNPs present in all accessions. Genetic relationships were estimated using neighbour-joining trees constructed with TASSEL 5.0 (Bradbury et al., 2007) software. The linkage disequilibrium (LD) between SNPs was evaluated using the squared Pearson's correlation coefficient ( $R^2$ ) and the LD heatmaps surrounding peaks in the GWAS were constructed using the "LDblockshow" R package (Dong et al., 2021).

## 2.7 | GWAS study

Two SNP genotype data sets, including the 404k CoreSNP Data set and the 6.5mio filtSNP Data set (v0.4) were downloaded from the OryzaSNP database (<http://snp-seek.irri.org/download.zul>) and used for GWAS. The 3K RG 6.5mio filtSNP Data set (v0.4) was used as association markers. After filtering out high missing rate sites ( $\geq 20\%$ ) and low minor allele frequency (MAF) sites ( $\leq 0.05$ ), a total of 4,102,565 (whole panel), 2,141,273 (*japonica* panel), and 2,982,786 (*indica* panel) SNPs were selected for analysis. Linear mixed model (LMM) was performed using Fast-LMM software (Lippert et al., 2011). Genetic relatedness was modelled as a random effect in LMM using the kinship matrix calculated from the 404k CoreSNP Data set. The genome-wide significance thresholds of the GWAS were determined using a modified Bonferroni correction as described (Li et al., 2012), in which the total number of SNPs for threshold calculation was replaced by the effective number of independent SNPs in this GWAS. Thresholds for controlling the Type I error rate were  $1.24 \times 10^{-6}$ ,  $3.06 \times 10^{-6}$  and  $1.42 \times 10^{-6}$  for the whole, *japonica* and *indica* panels, respectively. To obtain independent association signals, significant

SNPs were binned into peaks using a sliding window based on LD decay using the PLINK (Purcell et al., 2007) command `--clump-r2 0.6 --clump-kb 300 --clump-allow-overlap`. The distribution of independent significant SNP-glycerolipid associations across the genome was investigated by dividing the whole genome into 1 Mb partitions and the number of significant associations in each segment was counted. Plausible candidate genes for regulating variable glycerolipid content were selected based on: location < 150 kb to the most significant peak SNPs (Xie et al., 2015), annotated with, or biologically related to, traits and highly expressed in seeds.

## 2.8 | QTL mapping

ICIMapping 4.0 software was used to detect QTL by constructing a Bin-map in which the genotype of 9311 and Nipponbare was assigned as 2 and 0, respectively (Meng et al., 2015). The QTL were determined by a likelihood ratio test based on single-point analysis. QTL with logarithm of the odds ratio (LODs)  $\geq 3.0$  were considered as real. The allelic effect was calculated as the mean effect of replacing Nipponbare alleles with 9311 alleles at the QTL.

## 2.9 | Transgenic analysis

waxy mutants in two different genetic backgrounds, Wuyunjing 7 (9522) and Xiushui134 (XS134), were developed in a previous study (Meng et al., 2015). For OsLP1 overexpression, the coding sequences of OsLP1 were amplified from leaf cDNA of Nipponbare and 9311, and PCR products were cloned into pU1300 (Wuhan Towin Biotechnology Co., Ltd), which contains the maize ubiquitin promoter upstream of the cloning site. Plasmids were introduced into Nipponbare callus using *Agrobacterium tumefaciens* (EHA105)-mediated transformation (Toki et al., 2006). Primary transformants ( $T_0$  generation) and their progenies were genotyped by transgene-specific primers to confirm positive transgenic plants.  $T_2$  plants were used for seed glycerolipid profiling. Sequences of relevant primers are listed in Supporting Information: Table 1.

## 2.10 | qRT-PCR reaction

Total RNA was isolated using TRIzol reagent (Thermo Fisher Scientific). Quality and concentration of total RNA were analysed using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). qRT-PCR was carried out using gene-specific primers and SYBR Premix ExTaq reagent (Takara) with an ABI 7500 real-time PCR system (Applied Biosystems) according to the manufacturer's instructions. The real-time PCR system (Bio-Rad) was used with the following programme: 95°C for 2 min, 40 cycles of two-step amplification (95°C for 5 s and 55°C for 35 s). qRT-PCR was performed with three biological replicates, each with three technical repeats. *OsACTIN* was used as the internal control gene and

expression levels were calculated using the comparative cycle threshold ( $\Delta\Delta C_t$ ) method to quantify the relative expression level of the target genes. Sequences of relevant primers are listed in Supporting Information: Table 2.

## 2.11 | GUS staining

A 2.26 kb DNA fragment upstream of the transcriptional start codon of *OsLP1* was amplified from genomic DNA and introduced into CAMBIA1301::GUS by In-Fusion (Clontech). The *pOsLP1::GUS* construct was transformed into Nipponbare callus using *A. tumefaciens*. GUS activity was determined by staining as described previously (Jefferson et al., 1987). Sequences of relevant primers are listed in Supporting Information: Table 1.

## 2.12 | Subcellular localization

Vectors containing the 35S::OsLP1-GFP, the endoplasmic reticulum (ER) marker 35S::HDEL-mCherry and plasma membrane marker 35S::CD3-1007-mCherry constructs were transformed into *Agrobacterium* strain GV3101 and infiltrated into leaves of 4-week-old *Nicotiana benthamiana* plants (Sparkes et al., 2006). Fluorescence signals were observed under confocal microscope (Leica TCS SP5). Green fluorescent protein (GFP) and red fluorescence were imaged at 488 nm (excitation)/520–580 nm (emission) and 514 nm (excitation)/640–750 nm (emission), respectively. Sequences of relevant primers are listed in Supporting Information: Table 1.

## 2.13 | In vitro PC:diacylglycerol (DAG) cholinephosphotransferase (PDCT) enzyme activity

Coding sequences of *OsLP1<sup>NIP</sup>* and *OsLP1<sup>9311</sup>* were codon-optimized for recombinant expression in *Saccharomyces cerevisiae* and synthesized (Shanghai Generay Biotech Co., Ltd). The coding sequences were subcloned into the pYES expression vector and resulting constructs were transformed into *S. cerevisiae* mutant strain YNL 130C (MATa *cpt1::KanMX ept1*; Dharmacon Inc., Horizon Discovery), which is deficient in cholinephosphotransferase activity (Morash et al., 1994) using the lithium acetate-mediated method (Gietz & Schiestl, 2007). The empty pYES vector was used as a negative control. Sequences of relevant primers are listed in Supporting Information: Table 1.

Yeast cells were grown in the induction medium containing 0.67% (w/v) yeast nitrogen base, 0.2% (w/v) SC-Ura, 2% (w/v) galactose and 1% (w/v) raffinose until mid-log phase, and were harvested and used to prepare microsomal fractions as described previously (Xu et al., 2020). Protein concentration was measured using the Bradford assay (Bio-Rad) with bovine serum albumin as standard (Bradford, 1976), and normalized to 50  $\mu\text{g}/75 \mu\text{l}$  as crude enzymes for activity assay.

In vitro PDCT assays were performed according to the procedures described previously with slight modification (Lu et al., 2009; Wickramaratna et al., 2020). In brief, 2.5 nmol of  $^{14}\text{C}$ -labelled PC-C16:0/C18:1 (sn-1-palmitoyl-sn-2- $^{14}\text{C}$ ) oleoyl-PC), 100 nmol of unlabelled PC-C16:0/C18:1 (sn-1-palmitoyl-sn-2-oleoyl-PC), and 100 nmol of DAG-18:1/18:1 (sn-1,2-diolein) were mixed and immediately dried under a stream of  $\text{N}_2$ . Dried lipids were re-suspended in 25  $\mu\text{l}$  of 4 $\times$  reaction buffer containing 200 mM 3-(N-morpholino) profane-sulfonic acid/NaOH (pH 7.5), 80 mM  $\text{MgCl}_2$ , and 1.8% Triton X-100 in a sonication bath. The reaction was initiated with the addition of 75  $\mu\text{l}$  of microsomal fractions containing 50  $\mu\text{g}$  crude enzymes and incubated at 30°C for 10 min with shaking, followed by quenching with 100  $\mu\text{l}$  of chloroform/methanol (2:1, v/v). After centrifugation at 10,000g for 2 min, the bottom layer (lipid fraction) of the quenched reaction mixture was separated by thin-layer chromatography (TLC; 0.25 mm silica gel, DC-Fertigplatten, Macherey-Nagel) with hexane/diethyl ether/acetic acid (70:30:1.5, v/v). The TLC plate was exposed to a storage phosphor screen overnight and the resolved lipids were visualized by phosphor-imaging (Typhoon Trio Variable Mode Imager, GE Healthcare). The corresponding DAG spots were scraped from the plates and quantified for radioactivity using an LS 6500 multipurpose scintillation counter (Beckman-Coulter).

For the PC specificity assay, 2.5 nmol of  $^{14}\text{C}$ -labelled PC-16:0/16:0 (L- $\alpha$ -dipalmitoyl [2-palmitoyl-1- $^{14}\text{C}$ ]), PC-18:1/18:1 (L- $\alpha$ -dioleoyl [2-oleoyl-1- $^{14}\text{C}$ ]) or PC-16:1/18:1 (L- $\alpha$ -1-palmitoyl-2-oleoyl [2-oleoyl-1- $^{14}\text{C}$ ]), 100 nmol of corresponding unlabelled PC, and 100 nmol of DAG-18:1/18:1 (sn-1,2-diolein) were used in the assay. For the DAG specificity assay, 100 nmol of DAG-18:1/18:1 (sn-1,2-diolein), DAG-16:0/16:0 (sn-1,2-dipalmitin) or DAG-18:2/18:2 (sn-1,2-dilinolein), 2.5 nmol of  $^{14}\text{C}$ -labelled PC-18:1/18:1 (L- $\alpha$ -dioleoyl [2-oleoyl-1- $^{14}\text{C}$ ]), or PC-16:1/18:1 (L- $\alpha$ -1-palmitoyl-2-oleoyl [2-oleoyl-1- $^{14}\text{C}$ ]), and 100 nmol of corresponding unlabelled PC were used in the assay. All  $^{14}\text{C}$ -labelled PCs or DAGs were purchased from American Radiolabeled Chemicals.

## 2.14 | Haplotype network analysis

Genotype data were downloaded from the Rice SNP-Seek Database (Alexandrov et al., 2015) (<http://snp-seek.irri.org/>) and ECOGEMS database (Yao et al., 2019) (<http://150.109.59.144:3838/ECOGEMS/>). Available polymorphic sites with MAF > 0.05 in the promoter region and coding region of *OsLP1* were used to construct the haplotype network of *OsLP1*. An online haplotype viewer (<http://www.cibiv.at/%7Egreg/hapviewer>) was used to visualize the haplotype network.

## 2.15 | Comparative GWAS analysis

Flooding-related phenotype data (total internode length 7 days after submergence at 10 leaf stage) and genotype data were downloaded from a previous study (Kuroha et al., 2018). To allow comparison

between the flooding-related and glycerolipid traits, we used Fast-LMM software (Lippert et al., 2011) on the flooding GWAS data using our statistical GWAS model.

## 2.16 | Heading date measurement

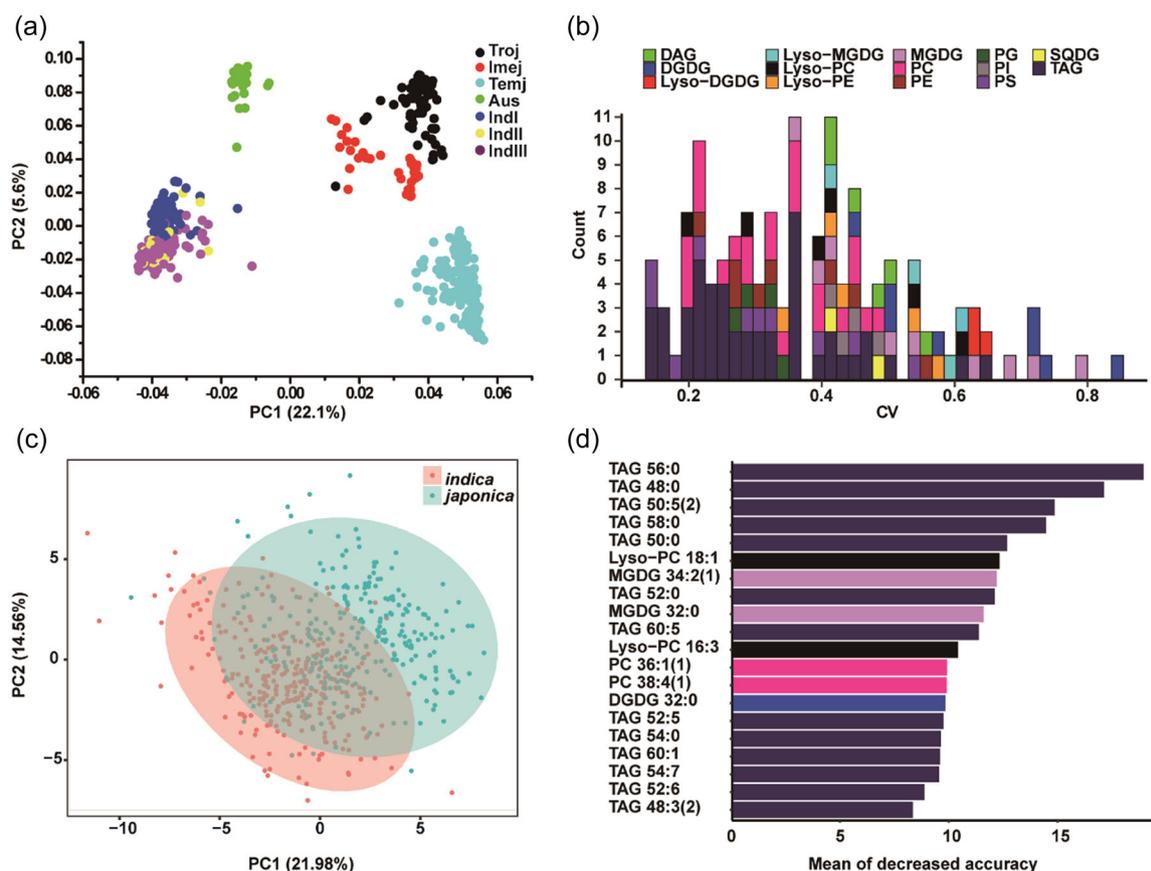
*OsLP1* overexpression plants ( $T_2$  generation), Nipponbare and CSSL49 plants were grown in the paddy field of Shanghai Jiao Tong University (31.03°N, 121.45°E) during the summer season of 2020. Heading date was recorded as the number of days from sowing to emergence of the first panicle.

## 3 | RESULTS

### 3.1 | Glycerolipid profiling in rice seeds

Two populations were used for glycerolipid profiling. The first contained 587 Asian cultivated rice accessions from the 3K Rice Genomes Project (<http://snp-seek.irri.org>; Alexandrov et al., 2015), including seven subspecies: *Aus*, *indical*, *indicall*, *indicalll*, *temperate*

*japonica*, *intermediate japonica*, and *tropical japonica* (Figure 1a and Supporting Information: Figure 1A,B and Data Set 1). The second population comprised 103 CSSLs developed from Nipponbare (*japonica*) as the recurrent parent and 9311 (*indica*) as the donor parent (Xu et al., 2017). Glycerolipid profiling using ultra performance UPLC-FT-MS (Hummel et al., 2011) identified 139 seed glycerolipids across the 587 accessions in one growing season (Supporting Information: Data Set 3), and 107 and 121 seed glycerolipids in the CSSL population grown over two seasons (Supporting Information: Data Sets 4 and 5). These glycerolipids included 14 categories: sulfoquinovosyl DAG, monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol (DGDG), lyso-MGDG, lyso-DGDG, DAG, TAG, PC, phosphatidylethanolamine (PE), lyso-PC, lyso-PE, phosphatidylinositol, phosphatidylserine and PG (Supporting Information: Data Set 2). All glycerolipids displayed substantial variation in their relative abundances (Supporting Information: Figure 2A-C and Data Set 3-5). Across the 587 accessions, a high degree of correlation was observed among glycerolipids within the same and between different categories (Supporting Information: Figure 2D). CVs of seed glycerolipids in 587 accessions ranged from 14.0% to 84.1% (Figure 1b), much lower than reported CVs of flavonoids that ranged from 71% to 1165%, but comparable to those of amino acids ranging from 58% to 106%



**FIGURE 1** Glycerolipid diversity in Asian cultivated rice. (a) Principal component analysis (PCA) of the genetic variance in 587 rice accessions. (b) Distribution of the genetic coefficients of variation (CVs) of 139 detected glycerolipids in 587 rice accessions. (c) PCA analysis of the detected 139 glycerolipids in 587 rice accessions. (d) The top 20 glycerolipids distinguishing *indica* from *japonica* accessions identified by random-forest analysis ( $n = 564$ , 301 *indica* and 262 *japonica*). [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

(Chen et al., 2016). Galactolipids such as MGDG and DGDG, generally had higher CVs (35%–84%, average 59%) than phospholipids, such as PC and lyso-PC (15%–62%, average 35%) and TAGs (14%–66%, average 30%; Figure 1b), supporting a stricter regulation of phospholipid (Colin & Jaillais, 2020) and TAG homeostasis (Xu & Shanklin, 2016) in rice.

PCA using all detected glycerolipids demonstrated that the first two components can separate *japonica* from *indica* cultivars with some overlap (Figure 1c), suggesting a role for seed glycerolipids in *indica*–*japonica* differentiation. The top 20 glycerolipids revealed by random-forest analysis (Hu et al., 2014) could discriminate *japonica* from *indica* varieties with >85% accuracy (Figure 1d and Supporting Information: Figure 3A). Among them, a series of saturated TAGs (from TAG 48:0 to TAG 58:0) were significantly higher and two lyso-PCs (lyso-PC 18:1 and lyso-PC 16:3) were significantly lower in *indica* seeds (Supporting Information: Figure 3B).

### 3.2 | Genetic basis of glycerolipid diversity in rice seeds

One-way ANOVA using data from CSSL populations grown over two seasons revealed that most glycerolipids displayed a high broad-sense heritability ( $H^2 > 0.4$ ; Figure 2A), showing that rice seed glycerolipids are mainly controlled by genetic determinants as proposed (Chen et al., 2016). To dissect the genetic bases underlying natural variation of glycerolipid metabolism in rice seeds, mGWAS was performed in the panel of 587 Asian cultivated rice accessions, which detected 1418 SNP–glycerolipid associations in the whole panel, 1098 associations in *indica* lines, and 1684 associations in *japonica* lines (Figure 2b and Supporting Information: Data Sets 6–8), indicating a high and heterogeneous level of genetic control of the seed glycerolipidome. Among these SNP–glycerolipid associations, likely candidate genes responsible for variable glycerolipids were searched based on: (1) location  $\leq 150$  kb to the most significant peak SNPs; (2) annotation with or biologically related to associated lipid traits; and (3) relatively high expression in seed. In total, 23 putative candidate genes were assigned to corresponding glycerolipid mQTL from the whole panel (Supporting Information: Table 2), including four known lipid metabolic genes: *PLD $\alpha$ 1* (Suzuki, 2011), *HMS1* (Chen, Zang, et al., 2020; Chen, Norton, et al., 2020), *WDA1* (Jung et al., 2006) and *HSD1* (Zhang et al., 2016). Considering that GWAS has limited power to identify rare alleles and linkage-based mapping is usually employed to complement GWAS in a single study (Lou et al., 2015), we additionally performed parallel linkage-based mapping using the 103 CSSL lines and revealed 55 mQTL in two seasons with LOD values  $> 3.0$ , explaining 6.5%–55.3% of the total phenotypic variation in 11 distinct lipid categories (Figure 2c and Supporting Information: Data Set 9). Notably, among these 55 mQTL, only one mQTL (Loci 34) could be repeatedly detected in two seasons (Supporting Information: Data Set 9), suggesting that genetic architecture of rice seed lipidome are also strongly affected by gene–environment interactions.

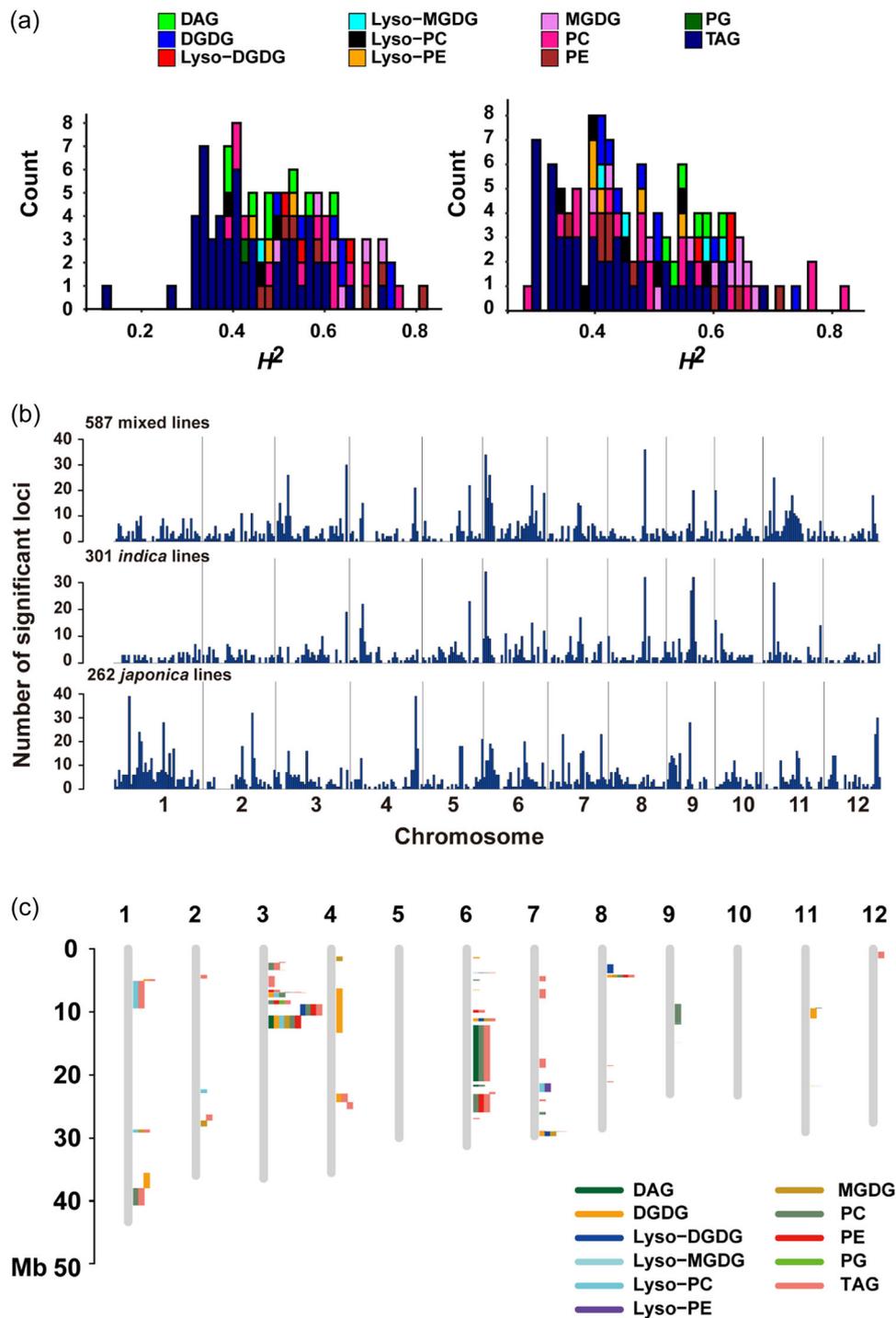
### 3.3 | Natural variation of *OsLP1* controls rice seed saturated TAG levels

To further explore the genetic basis of glycerolipid diversity between *indica* and *japonica* subspecies, we focused on two mQTL controlling saturated TAGs and lyso-PCs, respectively. An mQTL controlling levels of seed saturated TAGs (TAG 48:0–TAG 58:0) was repeatedly identified from CSSL populations grown over two seasons. This mQTL, designated *OsLP1*, was mapped to region 23.01–25.89 Mb on chromosome 6, which explained 22%–41% of the phenotypic variance (Supporting Information: Data Set 9). Glycerolipid profiling in seeds of CSSL49, a CSSL harbouring the *OsLP1*<sup>9311</sup> introgression segment in Nipponbare (Supporting Information: Figure 4A), revealed significantly higher levels of six saturated TAGs compared with Nipponbare seeds (Supporting Information: Figure 4B), confirming the function of *OsLP1* in saturated TAG accumulation.

Accordingly, mGWAS signals for these saturated TAGs (TAG 48:0–TAG 58:0) colocalized in the *OsLP1* region of chromosome 6; the most significant mGWAS signal (for TAG 52:0) was  $\sim 10$  kb upstream of the translation initiation site of *LOC\_Os06g40500/OsLP1* (Figure 3a and Supporting Information: Figure 5), encoding a PDCT. Its *Arabidopsis* orthologue, AtROD1, catalyses exchange of the PC head group between PC and DAG during seed TAG biosynthesis (Lu et al., 2009). Gene-based association analysis revealed that the most significant SNP associated with saturated TAGs in the *OsLP1* genomic region was a G/A missense mutation that caused a glutamate to lysine substitution (E57K; Figure 3b). Because CSS49 (*OsLP1*<sup>9311</sup>) seeds accumulated more saturated TAG than Nipponbare, we sequenced the coding region of *LOC\_Os06g40500* in 9311 and Nipponbare, and found the same G/A missense mutation (E57K) and an additional adjacent G/A SNP (G60S) in the coding regions of both lines (Figure 3c), suggesting a functional role for these two amino acids in the divergence of saturated TAGs between *japonica* and *indica* subspecies. Indeed, levels of saturated TAGs in seeds of transgenic lines overexpressing *OsLP1*<sup>9311</sup> were significantly higher than those in seeds of transgenic lines overexpressing *OsLP1*<sup>NIP</sup> in Nipponbare (Figure 3d and Supporting Information: Figure 6). This result provided further evidence that *OsLP1* regulates levels of saturated TAGs in rice seeds and the difference in seed saturated TAGs content between C SSL49 and Nipponbare was determined by divergent *OsLP1* function.

### 3.4 | Variation in *OsLP1* alters in vitro enzyme substrate preference

Although *OsLP1* is ubiquitously expressed in both vegetative and reproductive tissues with higher expression in leaves and mature seeds (Figure 3e), its expression levels were not significantly different between Nipponbare and CSSL49 lines (Figure 3e). Therefore, we sought to biochemically characterize *OsLP1* function, which was shown to localize to the plasma membrane and ER (Supporting Information: Figure 7A). Notably, both *OsLP1*<sup>NIP</sup> and *OsLP1*<sup>9311</sup>



**FIGURE 2** Quantitative trait loci (QTL) mapping and metabolite-based genome-wide association analysis (mGWAS) for seed glycerolipids. (A) Distribution of broad-sense heritability ( $H^2$ ) for 107 and 121 glycerolipids detected in chromosome segment substitution lines (CSSLs) grown in Yangzhou (left) and Sanya (right), respectively. (b) Distribution of mGWAS signals across the rice genome identified in the whole, *indica*, and *japonica* panels. *IndI*, *indica I*; *IndII*, *indica II*; *IndIII*, *indica III*; *Imej*, *intermedia japonica*; *Temj*, *temperate japonica*; *Troj*, *tropical japonica*. (c) Chromosomal distribution of the mQTL identified in CSSL lines; each coloured line represents an mQTL. DAG, diacylglycerol; DGDG, digalactosyldiacylglycerol; Lyso-DGDG, lyso-digalactosyldiacylglycerol; Lyso-MGDG, lyso-monogalactosyldiacylglycerol; Lyso-PC, lyso-phosphatidylcholine; Lyso-PE, lyso-phosphatidylethanolamine; MGDG, monogalactosyldiacylglycerol; PC, phosphatidylcholine; PE, phosphatidylethanolamine; PG, phosphatidylglycerol; TAG, triacylglycerol. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

share eight highly conserved amino acid residues present in AtROD1 (Supporting Information: Figure 7B), which are considered to be essential for PDCT enzyme activity (Lu et al., 2009). To test whether OsLP1<sup>NIP</sup> and OsLP1<sup>9311</sup> display PDCT activity, sn-1-palmitoyl-sn-2-

[14 C] oleoyl-PC (16:0/18:1-PC) and diolein were incubated with yeast microsomal fractions for 30 min and the production of radiolabeled-DAG was detected using TLC. Both OsLP1<sup>NIP</sup> and OsLP1<sup>9311</sup> but not negative controls were able to convert PC to

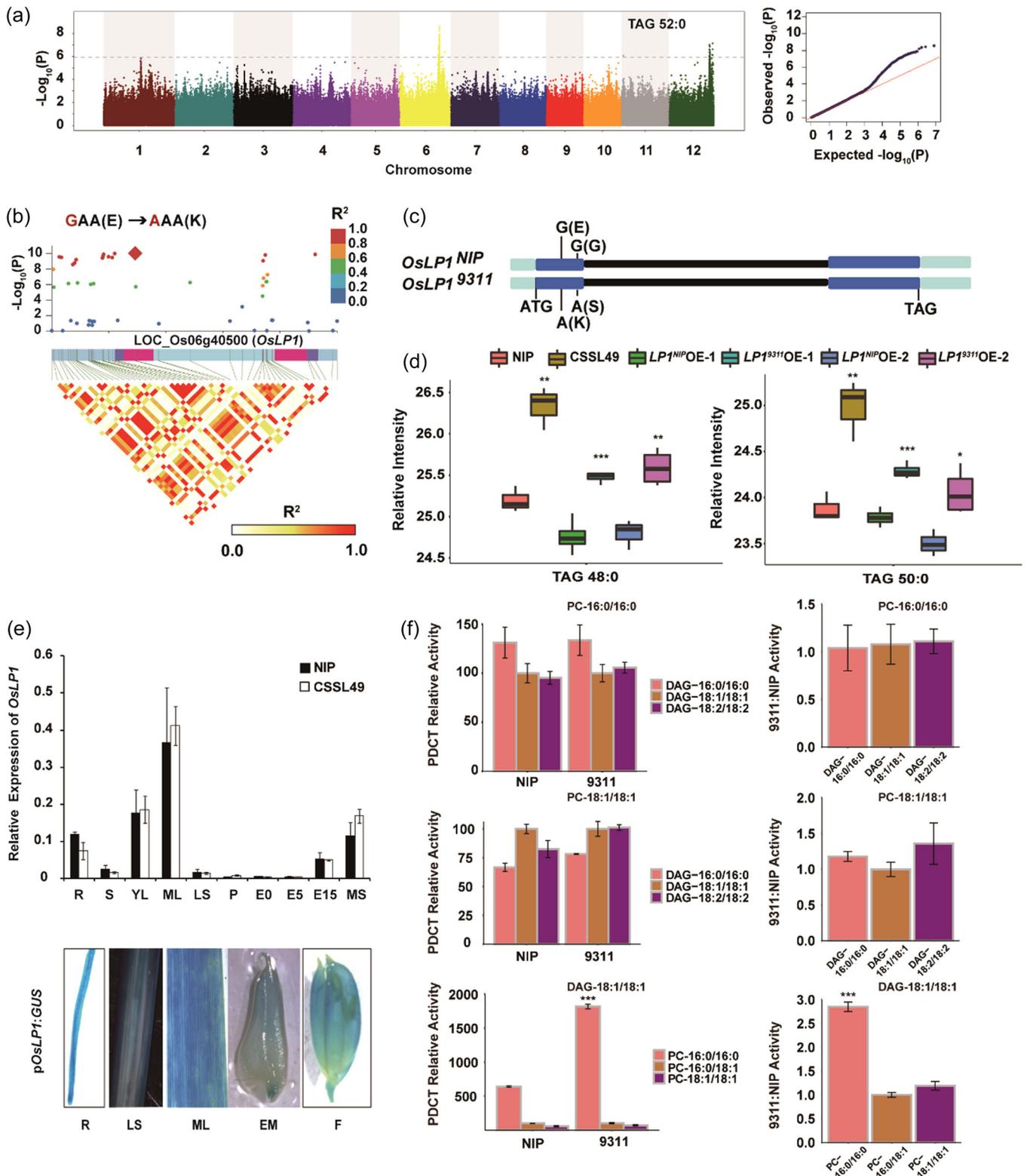


FIGURE 3 (See caption on next page)

DAG (Supporting Information: Figure 7C). We further conducted substrate specificity assays to investigate whether OsLP1<sup>NIP</sup> and OsLP1<sup>9311</sup> affect seed-saturated TAGs via different preferences towards substrates. When either <sup>14</sup>C-labelled PC-16:0/16:0 or PC-18:1/18:1 was used as the constant PC substrate, neither OsLP1 isoforms exhibited a preference toward DAG-16:0/16:0, DAG-18:1/18:1 or DAG-18:2/18:2 (Figure 3f). When <sup>14</sup>C-labelled DAG-18:1/18:1 was used as the constant DAG substrate, OsLP1<sup>NIP</sup> and OsLP1<sup>9311</sup> showed similar preferences for PC-16:0/18:1 or PC-18:1/18:1, but OsLP1<sup>9311</sup> displayed a threefold greater preference than OsLP1<sup>NIP</sup> for PC-16:0/16:0 (Figure 3f). These results suggest that the two amino acid substitutions (E57G and K60S) in OsLP1<sup>9311</sup> increase its substrate preference for PC with saturated fatty acid side chain such as PC-16:0/16:0, eventually leading to different accumulation profiles of saturated TAGs in *japonica* and *indica* seeds. Recognizing how minor sequence variations in *OsLP1* can alter enzyme substrate preference has deepened our understanding of PDCT activity in plant and provided a novel strategy to modify seed lipid saturation.

### 3.5 | Independent selection of *OsLP1* is associated with distinct ecological pressures

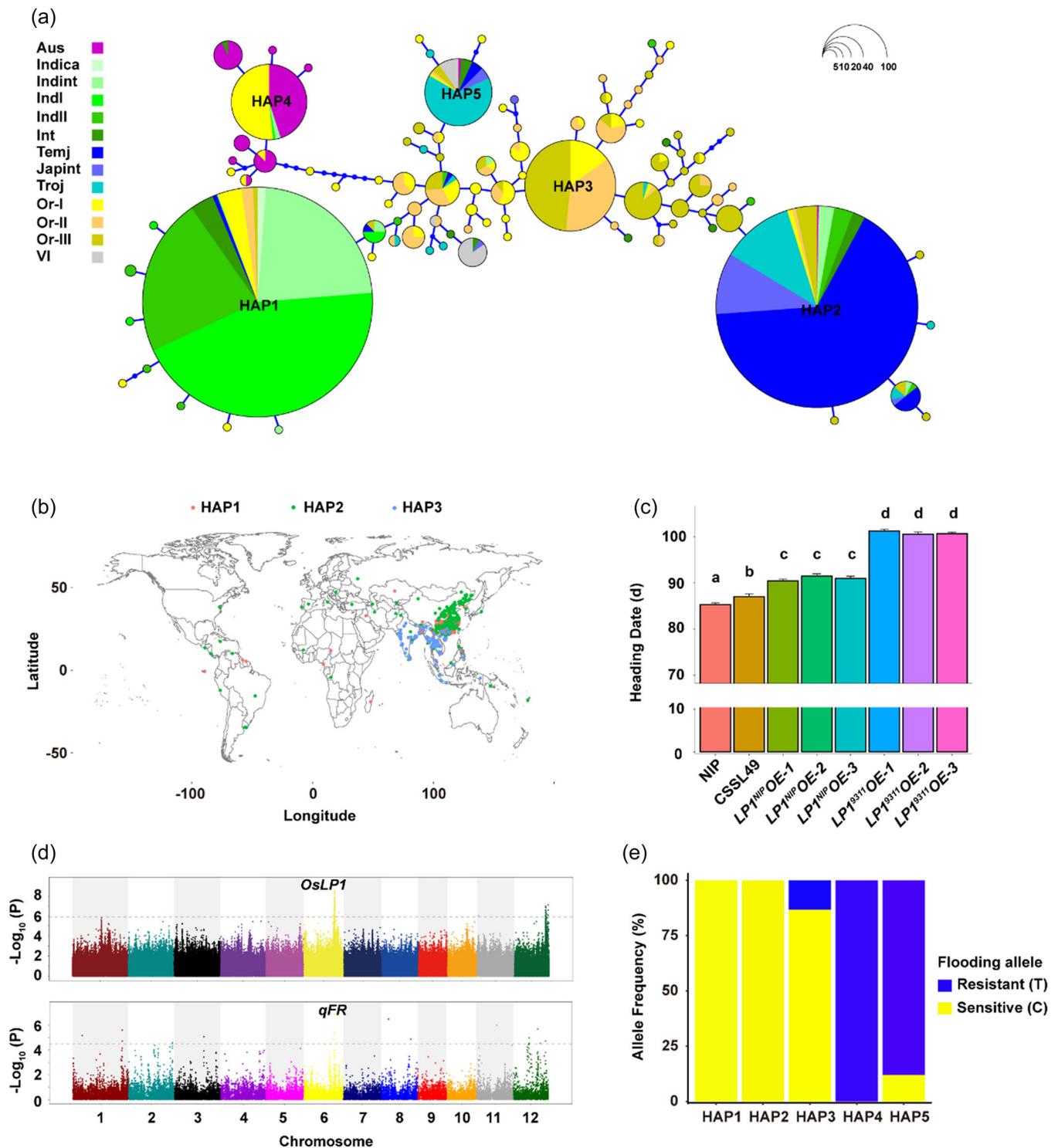
The diversification of metabolites within plant species has been intensively investigated (Beleggia et al., 2016; Chen et al., 2014; Wen et al., 2014); however, how diverse alleles arise to control metabolite diversity and how different ecological pressures shape these alleles remain largely unknown (Katz et al., 2021). To better understand whether the ecological pressures structured *OsLP1* alleles/haplotypes in Asian cultivated rice, we next explored the evolutionary transitional processes of *OsLP1*. Haplotype network analysis using SNPs in the *OsLP1* promoter and coding regions from 3024 Asian cultivated rice accessions (Alexandrov et al., 2015) revealed a dramatically unbalanced frequency distribution of *OsLP1* haplotypes across subspecies (Supporting Information: Figure 8A). Furthermore, focused analysis on the two causal SNPs between *OsLP1*<sup>NIP</sup> and

*OsLP1*<sup>9311</sup> (Figure 3c) revealed that *japonica* cultivars harbour two haplotypes (90.2% GG and 9.8% AG), while *indica* cultivars harbour three haplotypes (85.6% AA, 11.8% GG and 2.6% AG) with different frequencies. The AG haplotype, uncommon in *japonica* and *indica* cultivars, dominated in *Aus* (80.8%) and *Aro* (87.5%) cultivars (Supporting Information: Figure 8B), indicating that *OsLP1* contributes to the differentiation of distinct rice subspecies.

To further explore the origin and evolution of *OsLP1*, we performed additional haplotype network analysis using another 1612 cultivated (*O. sativa*) and 466 wild (*O. rufipogon*) rice accessions from ECOGEMS databases (Yao et al., 2019). We revealed a radial evolution of *OsLP1* haplotypes from an *O. rufipogon*-dominant ancestral haplotype (HAP3) to *O. sativa*-dominant haplotypes HAP1 (*OsLP1*<sup>9311</sup>), HAP2 (*OsLP1*<sup>NIP</sup>), HAP4 and HAP5, which were enriched in *indica*, *temperate japonica*, *Aus*, and *tropical japonica* accessions, respectively (Figure 4a). These results suggest that *OsLP1* haplotypes have been independently selected in each subspecies after their evolutions from the common ancestral haplotype.

Interestingly, the distinct geographic distribution of *O. rufipogon*-dominant *OsLP1* HAP3, *indica*-dominant *OsLP1* HAP1, and *japonica*-dominant *OsLP1* HAP2 revealed that enrichment of *OsLP1* haplotypes differs across subspecies with different latitudes (Figure 4b), suggesting the direct association of *OsLP1* differentiation with regional adaptation of rice. Considering that glycerolipids play important roles in the regulation of flowering (Nakamura et al., 2014, 2019; Susila et al., 2021) and constitutive expression pattern of *OsLP1* (Figure 3e), we investigated whether allelic variation between *indica* and *japonica* could affect heading date, an important adaptive feature of flowering plants. The heading date of CSSL49 introgression lines (HAP1; 9311 allele) in the paddy field was significantly delayed compared with Nipponbare lines (HAP2; Nipponbare allele), while the heading date of transgenic lines overexpressing *OsLP1*<sup>9311</sup> was delayed over 1 week compared with that of transgenic lines overexpressing *OsLP1*<sup>NIP</sup>, in the same Nipponbare genetic background, respectively (Figure 4c). Although the precise mechanism underlying flowering time regulation remains to be further investigated, our genetic evidence indicated that *OsLP1*

**FIGURE 3** Differences in *OsLP1* activity underlie natural variation in saturated triglyceride (TAG) content between *indica* and *japonica* seeds. (a) Manhattan (left, significant threshold  $P = 1.24 \times 10^{-6}$ ) and corresponding quantile–quantile (Q–Q) plot (right) of TAG 52:0 content. (b) Local Manhattan plot in the *OsLP1* genomic region (upper) and linkage disequilibrium between polymorphic sites in *OsLP1* (lower). The red diamond denotes the most significant missense variant that causes a glutamate to lysine replacement. (c) Schematic diagram of *OsLP1*<sup>9311</sup> and *OsLP1*<sup>NIP</sup> alleles. Exons are depicted by dark blue boxes, untranslated regions by light blue boxes, and introns by black lines. Missense mutations are indicated, along with the resulting amino acid substitution in brackets. (d) Effect of overexpression (OE) of *OsLP1*<sup>NIP</sup> and *OsLP1*<sup>9311</sup> in Nipponbare (NIP) on saturated TAG content. The horizontal bar within the box represents the median. The top and bottom of the box represent the 75th and 25th percentiles, respectively. The upper and lower whiskers extend to 1.5× the interquartile range, with outliers shown as black dots. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$  (Student's *t* test). (e) Expression analysis of *OsLP1* by qRT-PCR (upper, relative to *OsACTIN*) and GUS histochemistry (lower). EO, embryo at 0 days after flowering (DAF); E5, embryo at 5 DAF; E15, embryo at 15 DAF; EM, embryo; F, flower; LS, leaf sheath; ML, mature leaf; MS, mature seed; P, palea; R, root; S, shoot; YL, young leaf. (f) In vitro substrate specificity of Nipponbare and 9311 *OsLP1* isoforms. DAG specificity test using radio-labelled PC-16:0/16:0 or PC-18:1/18:1 as constant substrates (upper and middle, respectively); PC specificity test using radio-labelled DAG-18:1/18:1 as constant substrate (lower). Left panels, PDCT relative activity of each isoform; right panels, PDCT activity ratio between 9311 and Nipponbare isoforms. \*\*\* $p < 0.001$  (Student's *t* test). [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]



**FIGURE 4** Evolutionary analysis of *OsLP1* during rice domestication. (a) Haplotype network of *OsLP1* in wild (446 *O. rufipogon*) and cultivated (1612 *O. sativa*) rice accessions. Japint, *japonica intermedia*; Indint, *indica intermedia*; IndI, *indica I*; IndII, *indica II*; Int, *intermedia*, Or-I, II, III, *O. rufipogon I, II and III*, respectively; Temj, *temperate japonica*; Troj, *tropical japonica*; VI, others. (b) Geographical distribution of cultivated and wild rice with three main *OsLP1* haplotypes. (c) Heading date of Nipponbare (NIP), CSSL49 and transgenic plants overexpressing (OE) *OsLP1* alleles in the NIP background. Different letters indicate significant differences at  $p < 0.001$  (Duncan's multiple range test,  $n = 11-12$ ). (d) Manhattan plot showing the co-incidence of *OsLP1* (significant threshold  $P = 1.24 \times 10^{-6}$ ) with a reported flooding-resistant QTL (*qFR*, significant threshold  $P = 3.25 \times 10^{-5}$ ). (e) Allele frequency of the flooding-resistant genotype in rice accessions with different *OsLP1* haplotypes. [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

is a pleiotropic QTL contributing to seed saturated TAG levels and heading time. These results suggested that *OsLP1* enabled regional adaptation of subspecies during *indica-japonica* differentiation.

Notably, a specific *OsLP1* HAP4 haplotype was mainly found in *Aus* subspecies (Figure 4a). The *Aus* subspecies, originating mainly from Bangladesh and India, displays rich genetic diversity in abiotic stress resistance for local adaptation (Zhang et al., 2020). Interestingly, reanalysis of an existing flooding study (Kuroha et al., 2018) found that the most significant SNP-flooding resistant trait association (chr6: 24125294) colocalized with *OsLP1* (Figure 4d). Further investigation revealed that all *OsLP1* HAP4 accessions have the flooding-resistant allele, while *OsLP1* HAP1 and HAP2 accessions all have the flooding-sensitive allele *qFR<sup>S</sup>* (Figure 4e). The frequency of the flooding-resistant allele *qFR<sup>R</sup>* in ancestral *OsLP1* HAP3 varieties was lower than in *OsLP1* HAP4 (*Aus*) varieties (Figure 4e), suggesting that selection for improved flooding tolerance in *O. rufipogon* in flood-prone regions of southeast Asia fixed the *OsLP1* HAP4 in the resulting flooding-resistant *Aus* genotypes. Correspondingly, seeds of accessions with the flooding-resistant allele contained significantly higher levels of saturated TAGs than seeds of accessions with the flooding-sensitive allele (Supporting Information: Figure 8C). Above results suggest that during flooding stress, nature selection shaped seed glycerolipid profiles in local rice accessions due to collocation of *OsLP1* and flooding resistance QTL (*qFR*) in same genomic region.

### 3.6 | Natural variation of *Waxy* determines rice seed lyso-PC content

The Manhattan plot for lyso-PC 18:2 content revealed that accumulation of this glycerolipid in rice seed is largely determined by a single QTL on chromosome 6 (Figure 5a). Moreover, GWAS signals for lyso-PC 18:1 and lyso-PC 18:3 contents also mapped to this QTL (Supporting Information: Figure 9). To explore possible regulators associated with the variations in these three Lyso-PCs, we searched for the candidate genes near the genomic position of the most significant SNP. For lyso-PC 18:2 (Figure 5b), the most significant GWAS SNP (ID: 182785347) was about 78 kb away from the *Waxy* gene that directs amylose biosynthesis (Zhang et al., 2019). As lyso-PCs and amylose are known to form starch-lipid complexes in rice (Kusano et al., 2012), we speculated that *Waxy* gene is also associated with lyso-PCs content in rice seed. Next, we analysed lyso-PC levels in six different *Waxy* haplotypes in 531 Asian cultivated rice accessions and found that accessions harbouring *Waxy* HAP5 or *Waxy* HAP6 contained significantly lower levels of lyso-PC 18:2 than the other four haplotypes (Figure 5c). Notably, both *Waxy* HAP5 and *Waxy* HAP6 contained a 23 bp duplication in the second exon (Figure 5b), and the resulting frameshift mutation showed the most significant association with lyso-PC 18:2 content than any other SNPs (Figure 5d and Supporting Information: Figure 10). To confirm the function of this natural variation, we profiled glycerolipids in seeds of two available *waxy* mutants (Figure 5e) in two *japonica* cultivars (9522 and XS134) showing

improved glutinosity (Zhang et al., 2018) and found that levels of seed lyso-PCs in both mutants were significantly lower than those in wild type (Figure 5f). Our results revealed the pleiotropic effect of *Waxy* on lyso-PC and amylose contents, two important determinants of rice starch quality (Zhang et al., 2019).

## 4 | DISCUSSION

Asian cultivated rice has a wide geographic distribution across the world, containing a huge number of accessions that are rich in their phenotypic, physiological, and biochemical, as well as genomic variations. Elucidation of genetic and biochemical bases of metabolic variations across natural accessions not only provides better understanding of rice metabolism, but also reveals new strategies to improve yield, quality, and resilience to climate change. Recently, we have witnessed the successful application of mGWAS in rice to dissect genetic architecture of certain primary and secondary metabolites (Brotman et al., 2021; Chen et al., 2014, 2016; Dong et al., 2015; Peng et al., 2017; Sun et al., 2020; Zhou et al., 2021). Among them, only one study focused on the genetic basis of lipid-related oil biosynthesis, which identified four key genes regulating rice grain oil compositions using 11 detected oil components (Zhou et al., 2021). Even though glycerolipids have been shown to determine seed oil traits in soybean and maize (Liu et al., 2020; Wen et al., 2014), and to regulate flowering time in *Arabidopsis* and maize (Nakamura et al., 2014, 2019; Barnes et al., 2021), a systematic analysis of glycerolipid diversity and its underlying mechanisms in rice has not been attempted.

Here we have dissected the genetic base of glycerolipid metabolism in rice seed using lipidome-based GWAS and QTL, together with genetic, biochemical and bioinformatics approaches. We have identified more than 100 seed glycerolipids across the 587 Asian cultivated rice accessions and identified more than 1000 SNP-glycerolipid associations across the whole panel and in specific *indica* and *japonica* groups. These identified genetic loci can be used directly for marker-assisted breeding to create new rice varieties with targeted seed glycerolipid profiles (Zhou & Huang, 2019). Notably, similar to other primary and secondary metabolites (Chen et al., 2014, 2016), we observed distinct regulatory QTL/gene networks of seed glycerolipid in *indica* and *japonica* (Figure 2B). Combined with GWAS and bioinformatics analyses, 23 putative candidate genes for corresponding lipid metabolite traits were identified here. However, comprehensive further studies are needed to verify their causative roles in studied traits in future. Two key genes, *OsLP1* and *Waxy*, were shown to contribute to variable saturated TAG and lyso-PCs accumulations, respectively, in rice seeds. Our bioinformatics analysis also indicated that these two glycerolipid classes can be used as new biomarkers to distinguish *indica* from *japonica* varieties (Figure 1d and Supporting Information: Figure 3B).

Previously, it has been reported that domestication influenced the genetic and metabolic architecture of plant populations due to

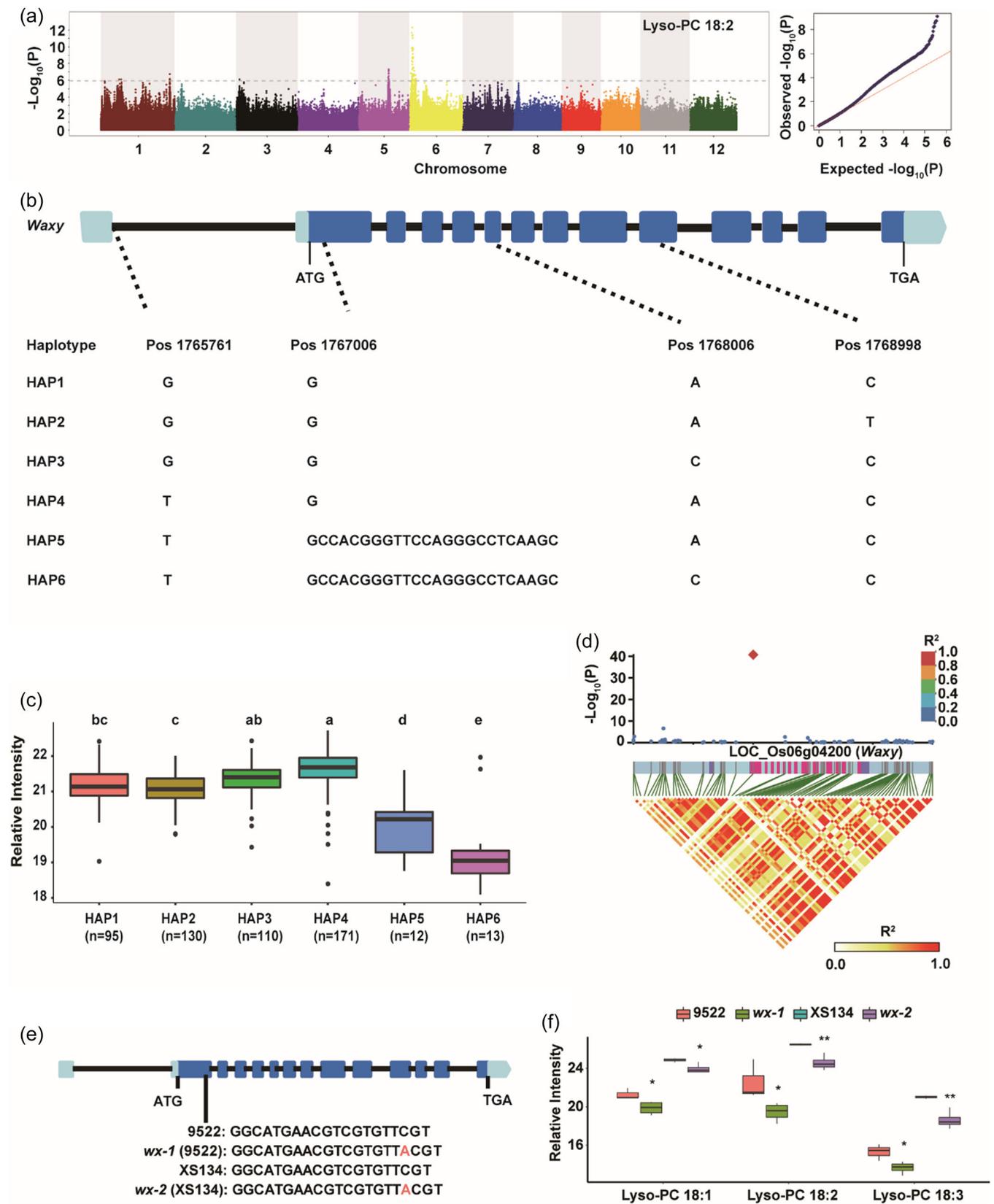
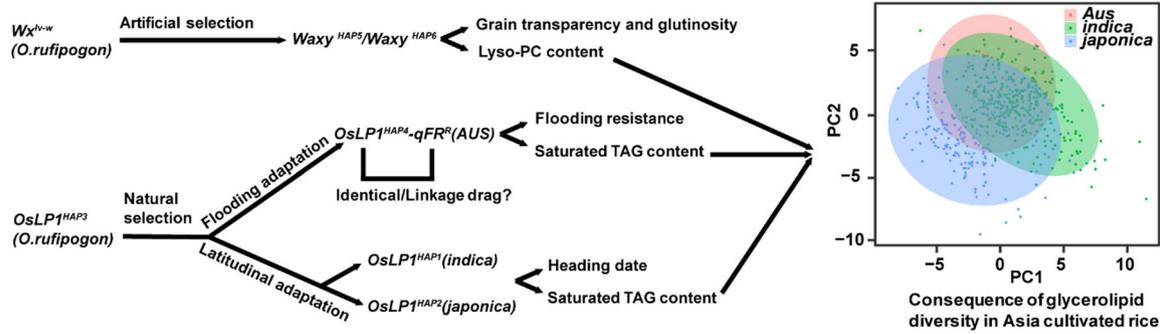


FIGURE 5 (See caption on next page)



**FIGURE 6** A proposed model for unintentional selection of seed glycerolipids during rice domestication.  $Wx^{lv-w}$  in wild rice (*O. rufipogon*) was deliberately selected for visible grain transparency and amylose content, but unintentionally for lyso-phosphatidylcholine (lyso-PC) content due to the pleiotropy of *Waxy* (upper). Different *OsLP1* alleles have undergone at least two independent natural selection events for local adaptation in different subspecies. *OsLP1*<sup>HAP4</sup> together with a flooding resistant QTL (*qFR*<sup>R</sup>) has been selected under deep water environments in *Aus* varieties, while *OsLP1* haplotypes that predominate in *indica* or *japonica* cultivars have been selected for adaptation to different growing latitudes, which links the saturated triglyceride (TAG) content with heading date phenotypes (lower). Variances in these two genes discriminate among three rice subspecies (principle component analysis [PCA] analysis, right). [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

local adaptation (Beleggia et al., 2016). For instance, domestication significantly altered the maize metabolome, and *Bxs*, *Pr1* (*F3'H*, flavonoid 3'-hydroxylase), *FHT1* (flavanone 3-hydroxylase1), and *ZmTPS1* (terpene synthase1) contribute to the metabolic divergence in benzoxazinoid, flavonoid and terpenoid pathway, respectively, between maize and its wild ancestor teosinte (Xu et al., 2019). Domestication also changed significantly the metabolome of wheat, particularly pathways for unsaturated fatty acids and amino acids (Beleggia et al., 2016). In almond, domestication targeted the *bHLH2* gene to reduce bitterness and the accumulation of cyanogenic diglucoside amygdalin (Sánchez-Pérez et al., 2019). In lettuce, domestication likely targeted three genes (*LG8749721*, *LG8763094* and *LG5482522*), contributing to observed different accumulations of galactinol, raffinose, quinate and chlorogenic acid (Zhang et al., 2020). In tomato, domestication selected five major loci and reduced steroidal glycoalkaloids in ripened fruits, rendering them more palatable (Zhu et al., 2018). This study demonstrated that domestication has significantly influenced rice seed glycerolipid profiles, and that both saturated TAG and lyso-PC contents have been unintentionally shaped.

*OsLP1* has been independently selected by at least two local adaptation events in different subspecies (Figure 6), similar to how rice *SD1* gene has been co-opted several times to optimize plant

growth in highly contrasting production systems (Kuroha et al., 2018). The first selection event was observed during *indica-japonica* differentiation, when allelic divergence of *OsLP1* established a link between glycerolipids and heading date (Figure 4c). Although molecular mechanisms underlying the time of transition from vegetative to reproductive growth have been extensively studied in rice (Guo et al., 2020), so far no heading date-related QTL has been associated with lipid metabolism. The discovered glycerolipid-heading date link highlights an alternative metabolic adaptation mechanism for different regional adaptation during *indica-japonica* differentiation and spread, supporting the essential conserved role of glycerolipids in flowering for both monocot and dicot plants (Barnes et al., 2021; Nakamura et al., 2014). The second *OsLP1* selection event occurred in *Aus* varieties via selection for flooding tolerance. Currently, we are not sure if *OsLP1* is the causal gene identical to the flooding-resistant QTL, but at least, our data indicated that the flooding-resistant QTL and *OsLP1* are extremely tightly linked, so that the *OsLP1* haplotype has been selected indirectly in *Aus* cultivars likely via linkage drag. A similar phenomenon has been observed in tomato, where selection for fruit size has led to substantial metabolomic changes through a hitchhiking effect (Zhu et al., 2018).

The selection of *Waxy* for lyso-PC content co-occurred with the artificial selection for amylose. Traditionally, people in southeast Asia

**FIGURE 5** *Waxy* is the causal gene for natural variation in lyso-PC content of rice seeds. (a) Manhattan plot (left, significant threshold  $P = 1.24 \times 10^{-6}$ ) and corresponding quantile-quantile (Q-Q) plot (right) of lyso-PC 18:2 content. (b) The genomic structure (upper) and haplotypes (lower) of the *Waxy* gene. Exons are depicted by dark blue boxes, untranslated regions by light blue boxes, and introns by black lines. Pos, position (bp) on chromosome 6. (c) Natural variation in lyso-PC 18:2 content across *Waxy* haplotypes. Different letters indicate significant differences at  $p < 0.001$  (Duncan's multiple range test). (d) Local Manhattan plot in the *Waxy* genomic region (upper) and linkage disequilibrium between polymorphic sites in *Waxy* (lower). The red diamond denotes the most significant 23 bp duplication variant. (e) Schematic diagram of mutations in the *Waxy* gene in two *japonica* cultivars (9522 and XS134), leading to the same 1 bp insertion causing premature protein truncation. (f) Contents of three lyso-PCs in wild type and *waxy* plants ( $n = 3-4$ ). The horizontal bar within the box represents the median. The top and bottom of the box represent the 75th and 25th percentiles, respectively. The upper and lower whiskers extend to  $1.5 \times$  the interquartile range, with outliers shown as black dots. \* $p < 0.05$ ; \*\* $p < 0.01$  (Student's *t* test). [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

prefer glutinous/low-amylose rice and have selected rice cultivars with the relevant *Waxy* haplotype/genotype (Yamanaka et al., 2004); this artificial selection for amylose content has indirectly affected lyso-PC content due to *Waxy* pleiotropy (Figure 6). Unintended selection of metabolic traits through pleiotropic QTL has also occurred in maize via the classic domestication gene *Tb1*, originally selected for branching (Studer et al., 2011) but later found to affect sugar metabolism (Dong et al., 2019), and likely occurs in other domesticated crops. Although the association of *Waxy* with lyso-PCs in rice seeds was previously reported (Gong et al., 2013), the function of *Waxy* in glycerolipid accumulation has not been experimentally verified until now (Figure 5f). A previous study revealed that lyso-PCs (16:0 and 18:2) are major starch lysophospholipids in rice grain, and that lyso-PCs and starch often combine together to form an amylose-lipid complex, significantly affecting starch qualities (Tong, 2016). Our data here suggest a role for *Waxy* in lyso PC-starch-complex formation and lipid metabolism in rice grain, contributing significantly to grain qualities. Knowledge of *Waxy* pleiotropy is valuable for molecular breeding: if both low amylose and low lyso-PC traits are desirable, *Waxy* HAP5 and *Waxy* HAP6 variants can be targeted directly; however, if only one trait is required, compensating loci/genes will need to be recruited (Chen & Lübberstedt, 2010).

Our study has explored the diversity of seed glycerolipids in Asian cultivated rice, highlighting an important role for genetic correlations (such as gene pleiotropy and linkage drag) in shaping glycerolipid profiles in different subspecies during rice domestication.

#### AUTHOR CONTRIBUTIONS

Jun Hong, Jianxin Shi and Dabing Zhang conceived and designed the study. Jun Hong, Yang Xu, Leah Rosental and Dawei Xu executed most experiments. Isabel Orf, Hui Zhang, Su Su, Shaoxing Bai, Mohammed Ashraf, Chaoyang Hu and Fengli Zhang assisted with experiments. Changquan Zhang, Zhijing Luo, Mingjiao Chen and Xiaofei Chen helped in plant transformation and field management. Wengsheng Wang, Zhikang Li, Jianlong Xu, Qiaoquan Liu, Hui Zhang and Wanqi Liang assisted with experimental design. Jun Hong, Leah Rosental, Yang Xu, Alisdair Fernie, Zhiqiang Hu, Guanqun Chen, Yariv Brotman, Jianxin Shi and Dabing Zhang analysed the data. Alisdair Fernie, Yariv Brotman, Dabing Zhang and Jianxin Shi supervised the experiments. Jun Hong, Alisdair Fernie, Natalie Betts, Yariv Brotman, Dabing Zhang and Jianxin Shi wrote the paper.

#### ACKNOWLEDGEMENT

This study was supported by National Natural Sciences Foundation of China (31971907, 31970803 and 31861163002) and the SJTU JiRLMDS Joint Research Fund (MDS-JF-2019B02).

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### ORCID

Mohammed Ashraf  <http://orcid.org/0000-0003-0862-6246>

Guanqun Chen  <http://orcid.org/0000-0001-5790-3903>

Jianxin Shi  <http://orcid.org/0000-0002-7717-0863>

#### REFERENCES

- Alexandrov, N., Tai, S., Wang, W., Mansueto, L., Palis, K., Fuentes, R.R. et al. (2015) SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Research*, 43(D1), D1023–D1027.
- Barnes, A.C., Rodríguez-Zapata, F., Blöcher-Juárez, K.A., Gates, D.J., Janzen, G.M., Kur, A. et al. (2021) Teosinte introgression modulates phosphatidylcholine levels and induces early maize flowering time. *bioRxiv*, 2021, Available at <https://doi.org/10.1101/2021.01.25.426574>
- Beleggia, R., Rau, D., Laidò, G., Platani, C., Nigro, F., Fragasso, M. et al. (2016) Evolutionary metabolomics reveals domestication-associated changes in tetraploid wheat kernels. *Molecular Biology and Evolution*, 33(7), 1740–1753.
- Bradbury, L.M.T., Fitzgerald, T.L., Henry, R.J., Jin, Q. & Waters, D.L.E. (2005) The gene for fragrance in rice. *Plant Biotechnology Journal*, 3(3), 363–370.
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. & Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633–2635.
- Bradford, M.M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(1), 248–254.
- Brotman, Y., Llorente-Wiegand, C., Oyong, G., Badoni, S., Misra, G., Anacleto, R. et al. (2021) The genetics underlying metabolic signatures in a brown rice diversity panel and their vital role in human nutrition. *The Plant Journal*, 106(2), 507–525.
- Chen, C., Norton, G.J. & Price, A.H. (2020) Genome-wide association mapping for salt tolerance of rice seedlings grown in hydroponic and soil systems using the Bengal and Assam Aus panel. *Frontiers in Plant Science*, 11, 576479.
- Chen, H., Zhang, Z., Ni, E., Lin, J., Peng, G., Huang, J. et al. (2020) HMS1 interacts with HMS11 to regulate very-long-chain fatty acid biosynthesis and the humidity-sensitive genic male sterility in rice (*Oryza sativa*). *New Phytologist*, 225(5), 2077–2093.
- Chen, Q., Li, W., Tan, L. & Tian, F. (2021) Harnessing knowledge from maize and rice domestication for new crop breeding. *Molecular Plant*, 14(1), 9–26.
- Chen, W., Gao, Y., Xie, W., Gong, L., Lu, K., Wang, W. et al. (2014) Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nature Genetics*, 46(7), 714–721.
- Chen, W., Wang, W., Peng, M., Gong, L., Gao, Y., Wan, J. et al. (2016) Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nature Communications*, 7(1), 12767.
- Chen, Y. & Lübberstedt, T. (2010) Molecular basis of trait correlations. *Trends in Plant Science*, 15(8), 454–461.
- Colin, L.A. & Jaillais, Y. (2020) Phospholipids across scales: lipid patterns and plant development. *Current Opinion in Plant Biology*, 53, 1–9.
- Dong, S.S., He, W.M., Ji, J.J., Zhang, C., Guo, Y. & Yang, T.L. (2021) LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings in Bioinformatics*, 22(4), bbaa227.
- Dong, X., Gao, Y., Chen, W., Wang, W., Gong, L., Liu, X. et al. (2015) Spatiotemporal distribution of phenolamides and the genetics of natural variation of hydroxycinnamoyl spermidine in rice. *Molecular Plant*, 8(1), 111–121.
- Dong, Z., Xiao, Y., Govindarajulu, R., Feil, R., Siddoway, M.L., Nielsen, T. et al. (2019) The regulatory landscape of a core maize domestication module controlling bud dormancy and growth repression. *Nature Communications*, 10(1), 3810.
- Fernie, A.R. & Yan, J. (2019) De novo domestication: an alternative route toward new crops for the future. *Molecular Plant*, 12(5), 615–631.
- Gietz, R.D. & Schiestl, R.H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nature Protocols*, 2(1), 31–34.

- Gong, L., Chen, W., Gao, Y., Liu, X., Zhang, H., Xu, C. et al. (2013) Genetic analysis of the metabolome exemplified using a rice population. *Proceedings of the National Academy of Sciences of the United States of America*, 110(50), 20320–20325.
- Guo, T., Mu, Q., Wang, J., Vanous, A.E., Onogi, A., Iwata, H. et al. (2020) Dynamic effects of interacting genes underlying rice flowering-time phenotypic plasticity and global adaptation. *Genome Research*, 30(5), 673–683.
- Gutaker, R.M., Groen, S.C., Bellis, E.S., Choi, J.Y., Pires, I.S., Bocinsky, R.K. et al. (2020) Genomic history and ecology of the geographic spread of rice. *Nature Plants*, 6(5), 492–502.
- Horn, P.J. & Benning, C. (2016) The plant lipidome in human and environmental health. *Science*, 353(6305), 1228–1232.
- Hu, C., Shi, J., Quan, S., Cui, B., Kleessen, S., Nikoloski, Z. et al. (2014) Metabolic variation between japonica and indica rice cultivars as revealed by non-targeted metabolomics. *Scientific Reports*, 4(1), 5067.
- Huang, X., Kurata, N., Wei, X., Wang, Z.-X., Wang, A., Zhao, Q. et al. (2012) A map of rice genome variation reveals the origin of cultivated rice. *Nature*, 490(7421), 497–501.
- Hummel, J., Segu, S., Li, Y., Irgang, S., Jueppner, J. & Giavalisco, P. (2011) Ultra-performance liquid chromatography and high-resolution mass spectrometry for the analysis of plant lipids. *Frontiers in Plant Science*, 2(54), 54.
- Izawa, T., Konishi, S., Shomura, A. & Yano, M. (2009) DNA changes tell us about rice domestication. *Current Opinion in Plant Biology*, 12(2), 185–192.
- Jefferson, R.A., Kavanagh, T.A. & Bevan, M.W. (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *The EMBO Journal*, 6(13), 3901–3907.
- Jung, K.-H., Han, M.-J., Lee, D.-y., Lee, Y.-S., Schreiber, L., Franke, R. et al. (2006) Wax-deficient *anther1* is involved in cuticle and wax production in rice anther walls and is required for pollen development. *The Plant Cell*, 18(11), 3015–3032.
- Katz, E., Li, J.-J., Jaegle, B., Ashkenazy, H., Abrahams, S.R., Bagaza, C. et al. (2021) Genetic variation, environment and demography intersect to shape *Arabidopsis* defense metabolite variation across Europe. *eLife*, 10, e67784.
- Kim, S.R., Torollo, G., Yoon, M.R., Kwak, J., Lee, C.K., Prahalada, G.D. et al. (2018) Loss-of-function alleles of Heading date 1 (*Hd1*) are associated with adaptation of temperate *Japonica* rice plants to the tropical region. *Frontiers in Plant Science*, 9, 1827.
- Kuroha, T., Nagai, K., Gamuyao, R., Wang, D.R., Furuta, T., Nakamori, M. et al. (2018) Ethylene-gibberellin signaling underlies adaptation of rice to periodic flooding. *Science*, 361(6398), 181–186.
- Kusano, M., Fukushima, A., Fujita, N., Okazaki, Y., Kobayashi, M., Oitome, N.F. et al. (2012) Deciphering starch quality of rice kernels using metabolite profiling and pedigree network analysis. *Molecular Plant*, 5(2), 442–451.
- Lapidot-Cohen, T.L., Rosental, Y. & Brotman, Y. (2020) Liquid chromatography–mass spectrometry-based analysis for lipophilic compound profiling in plants. *Current Protocols in Plant Biology*, 5(2), e20109. <https://doi.org/10.1002/cppb.20109>
- Lavell, A.A. & Benning, C. (2019) Cellular organization and regulation of plant glycerolipid metabolism. *Plant and Cell Physiology*, 60(6), 1176–1183.
- Li, H., Thrash, A., Tang, J.D., He, L., Yan, J. & Warburton, M.L. (2019) Leveraging GWAS data to identify metabolic pathways and networks involved in maize lipid biosynthesis. *The Plant Journal*, 98(5), 853–863.
- Li, M.-X., Yeung, J.M.Y., Cherny, S.S. & Sham, P.C. (2012) Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Human Genetics*, 131(5), 747–756.
- Lippert, C., Listgarten, J., Liu, Y., Kadie, C.M., Davidson, R.I. & Heckerman, D. (2011) FaST linear mixed models for genome-wide association studies. *Nature Methods*, 8(10), 833–835.
- Liu, C., Ou, S., Mao, B., Tang, J., Wang, W., Wang, H. et al. (2018) Early selection of bZIP73 facilitated adaptation of japonica rice to cold climates. *Nature Communications*, 9(1), 3302.
- Liu, J.-Y., Li, P., Zhang, Y.-W., Zuo, J.-F., Li, G., Han, X. et al. (2020) Three-dimensional genetic networks among seed oil-related traits, metabolites and genes reveal the genetic foundations of oil synthesis in soybean. *The Plant Journal*, 103(3), 1103–1124.
- Lou, Q., Chen, L., Mei, H., Wei, H., Feng, F., Wang, P. et al. (2015) Quantitative trait locus mapping of deep rooting by linkage and association analysis in rice. *Journal of Experimental Botany*, 66(15), 4749–4757.
- Lu, C., Xin, Z., Ren, Z., Miquel, M. & Browse, J. (2009) An enzyme regulating triacylglycerol composition is encoded by the *ROD1* gene of *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America*, 106(44), 18837–18842.
- Lyu, J., Huang, L., Zhang, S., Zhang, Y., He, W., Zeng, P. et al. (2020) Neofunctionalization of a Teosinte branched 1 homologue mediates adaptations of upland rice. *Nature Communications*, 11(1), 725.
- Matsuda, F., Okazaki, Y., Oikawa, A., Kusano, M., Nakabayashi, R., Kikuchi, J. et al. (2012) Dissection of genotype–phenotype associations in rice grains using metabolome quantitative trait loci analysis. *The Plant Journal*, 70(4), 624–636.
- Meng, L., Li, H., Zhang, L. & Wang, J. (2015) QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *The Crop Journal*, 3(3), 269–283.
- Morash, S.C., McMaster, C.R., Hjelmstad, R.H. & Bell, R.M. (1994) Studies employing *Saccharomyces cerevisiae* *cpt1* and *ept1* null mutants implicate the *CPT1* gene in coordinate regulation of phospholipid biosynthesis. *Journal of Biological Chemistry*, 269(46), 28769–28776.
- Nakamura, Y., Andrés, F., Kanehara, K., Liu, Y.C., Dörmann, P. & Coupland, G. (2014) *Arabidopsis* florigen *FT* binds to diurnally oscillating phospholipids that accelerate flowering. *Nature Communications*, 5(1), 3553.
- Nakamura, Y., Lin, Y.C., Watanabe, S., Liu, Y.C., Katsuyama, K., Kanehara, K. et al. (2019) High-resolution crystal structure of *Arabidopsis* FLOWERING LOCUS T illuminates its phospholipid-binding site in flowering. *iScience*, 21, 577–586.
- Peng, M., Shahzad, R., Gul, A., Subthain, H., Shen, S., Lei, L. et al. (2017) Differentially evolved glucosyltransferases determine natural variation of rice flavone accumulation and UV-tolerance. *Nature Communications*, 8(1), 1975.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D. et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81(3), 559–575.
- Raj, A., Stephens, M. & Pritchard, J.K. (2014) fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics*, 197(2), 573–589.
- Sánchez-Pérez, R., Pavan, S., Mazzeo, R., Moldovan, C., Aiese Cigliano, R., Del Cuento, J. et al. (2019) Mutation of a bHLH transcription factor allowed almond domestication. *Science*, 364(6445), 1095–1098.
- Shomura, A., Izawa, T., Ebana, K., Ebitani, T., Kanegae, H., Konishi, S. et al. (2008) Deletion in a gene associated with grain size increased yields during rice domestication. *Nature Genetics*, 40(8), 1023–1028.
- Sparkes, I.A., Runions, J., Kearns, A. & Hawes, C. (2006) Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nature Protocols*, 1(4), 2019–2025.
- Studer, A., Zhao, Q., Ross-Ibarra, J. & Doebley, J. (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nature Genetics*, 43(11), 1160–1163.

- Sun, Y., Shi, Y., Liu, G., Yao, F., Zhang, Y., Yang, C. et al. (2020) Natural variation in the OsbZIP18 promoter contributes to branched-chain amino acid levels in rice. *New Phytologist*, 228(5), 1548–1558.
- Susila, H., Jurić, S., Liu, L., Gawarecka, K., Chung, K.S., Jin, S. et al. (2021) Florigen sequestration in cellular membranes modulates temperature-responsive flowering. *Science*, 373(6559), 1137–1142.
- Suzuki, Y. (2011) Isolation and characterization of a rice (*Oryza sativa* L.) mutant deficient in seed phospholipase D, an enzyme involved in the degradation of oil-body membranes. *Crop Science*, 51(2), 567–573.
- Toki, S., Hara, N., Ono, K., Onodera, H., Tagiri, A., Oka, S. et al. (2006) Early infection of scutellum tissue with *Agrobacterium* allows high-speed transformation of rice. *The Plant Journal*, 47(6), 969–976.
- Tong, C. (2016) Lysophospholipids in rice (*Oryza sativa* L.): identification, genetic diversity and their relations to grain qualities. PhD thesis. Lismore, NSW: Southern Cross University.
- Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z. et al. (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature*, 557(7703), 43–49.
- Welti, R. & Wang, X. (2004) Lipid species profiling: a high-throughput approach to identify lipid compositional changes and determine the function of genes involved in lipid metabolism and signaling. *Current Opinion in Plant Biology*, 7(3), 337–344.
- Wen, W., Brotman, Y., Willmitzer, L., Yan, J. & Fernie, A.R. (2016) Broadening our portfolio in the genetic improvement of maize chemical composition. *Trends in Genetics*, 32(8), 459–469.
- Wen, W., Li, D., Li, X., Gao, Y., Li, W., Li, H. et al. (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nature Communications*, 5(1), 3438.
- Wickramaratna, A.D., Siloto, R., Mietkiewska, E., Singer, S.D. & Weselake, R.J. (2015) Heterologous expression of flax PHOSPHOLIPID: DIACYLGLYCEROL CHOLINEPHOSPHOTRANSFERASE (PDCT) increases polyunsaturated fatty acid content in yeast and *Arabidopsis* seeds. *BMC Biotechnology*, 15, 63. <https://doi.org/10.1186/s12896-015-0156-6>
- Xie, W., Wang, G., Yuan, M., Yao, W., Lyu, K., Zhao, H. et al. (2015) Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. *Proceedings of the National Academy of Sciences of the United States of America*, 112(39), E5411–E5419.
- Xu, C. & Shanklin, J. (2016) Triacylglycerol metabolism, function, and accumulation in plant vegetative tissues. *Annual Review of Plant Biology*, 67(1), 179–206.
- Xu, G., Cao, J., Wang, X., Chen, Q., Jin, W., Li, Z. et al. (2019) Evolutionary metabolomics identifies substantial metabolic divergence between maize and its wild ancestor, *Teosinte*. *The Plant Cell*, 31(9), 1990–2009.
- Xu, Y., Caldo, K.M.P., Falarz, L., Jayawardhane, K. & Chen, G. (2020) Kinetic improvement of an algal diacylglycerol acyltransferase 1 via fusion with an acyl-CoA binding protein. *The Plant Journal*, 102(4), 856–871.
- Xu, Z., Li, S., Zhang, C., Zhang, B., Zhu, K., Zhou, Y. et al. (2017) Genetic connection between cell-wall composition and grain yield via parallel QTL analysis in indica and japonica subspecies. *Scientific Reports*, 7(1), 12561.
- Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L. et al. (2008) Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. *Nature Genetics*, 40(6), 761–767.
- Yamanaka, S., Nakamura, I., Watanabe, K.N. & Sato, Y.I. (2004) Identification of SNPs in the waxy gene among glutinous rice cultivars and their evolutionary significance during the domestication process of rice. *Theoretical and Applied Genetics*, 108(7), 1200–1204.
- Yang, J., Lee, S.H., Goddard, M.E. & Visscher, P.M. (2011) GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88(1), 76–82.
- Yao, W., Huang, F., Zhang, X. & Tang, J. (2019) ECOGEMS: efficient compression and retrieve of SNP data of 2058 rice accessions with integer sparse matrices. *Bioinformatics*, 35(20), 4181–4183.
- Zhan, C., Lei, L., Liu, Z., Zhou, S., Yang, C., Zhu, X. et al. (2020) Selection of a subspecies-specific diterpene gene cluster implicated in rice disease resistance. *Nature Plants*, 6(12), 1447–1454.
- Zhang, C., Zhu, J., Chen, S., Fan, X., Li, Q., Lu, Y. et al. (2019) *Wxlv*, the ancestral allele of rice waxy gene. *Molecular Plant*, 12(8), 1157–1166.
- Zhang, J., Zhang, H., Botella, J.R. & Zhu, J.-K. (2018) Generation of new glutinous rice by CRISPR/Cas9-targeted mutagenesis of the waxy gene in elite rice varieties. *Journal of Integrative Plant Biology*, 60(5), 369–375.
- Zhang, W., Alseekh, S., Zhu, X., Zhang, Q., Fernie, A.R., Kuang, H. et al. (2020) Dissection of the domestication-shaped genetic architecture of lettuce primary metabolism. *The Plant Journal*, 104(3), 613–630.
- Zhang, Z., Cheng, Z., Gan, L., Zhang, H., Wu, F., Lin, Q. et al. (2016) OsHSD1, a hydroxysteroid dehydrogenase, is involved in cuticle formation and lipid homeostasis in rice. *Plant Science*, 249, 35–45.
- Zhou, H., Xia, D., Li, P., Ao, Y., Xu, X., Wan, S. et al. (2021) Genetic architecture and key genes controlling the diversity of oil composition in rice grains. *Molecular Plant*, 14(3), 456–469.
- Zhou, X. & Huang, X. (2019) Genome-wide association studies in rice: how to solve the low power problems? *Molecular Plant*, 12(1), 10–12.
- Zhu, G., Wang, S., Huang, Z., Zhang, S., Liao, Q., Zhang, C. et al. (2018) Rewiring of the fruit metabolome in tomato breeding. *Cell*, 172(1), 249–261.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Hong, J., Rosental, L., Xu, Y., Xu, D., Orf, I., Wang, W. et al. (2022) Genetic architecture of seed glycerolipids in Asian cultivated rice. *Plant, Cell & Environment*, 1–17. <https://doi.org/10.1111/pce.14378>